THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 38-08

WOLFGANG POLASEK
Institute for Advanced Studies, Austria
and The Rimini Centre of Economics Analisys, Italy

RICHARD SELLNER
Institute for Advanced Studies, Austria
and The Rimini Centre of Economics Analisys

"SPATIAL CHOW-LIN METHODS: BAYESIAN AND ML FORECAST COMPARISONS"

The Rimini Centre for Economic Analysis (RCEA) was established in March 2007. RCEA is a private, non-profit organization dedicated to independent research in Applied and Theoretical Economics and related fields. RCEA organizes seminars and workshops, sponsors a general interest journal *The Review of Economic Analysis*, and organizes a biennial conference: *Small Open Economies in the Globalized World* (SOEGW). Scientific work contributed by the RCEA Scholars is published in the RCEA Working Papers series.

The Rimini Centre for Economic Analysis
Legal address: Via Angherà, 22 – Head office: Via Patara, 3 - 47900 Rimini (RN) – Italy
www.rcfea.org - secretary@rcfea.org

SPATIAL CHOW-LIN METHODS: BAYESIAN AND ML FORECAST COMPARISONS

Wolfgang Polasek [a] Richard Sellner [b]

[a] *Institute for Advanced Studies, Stumpergasse 56, 1060 Vienna*

[b] *Institute for Advanced Studies, Stumpergasse 56, 1060 Vienna*

Abstract

Completing data that are collected in disaggregated and heterogeneous spatial units is a quite frequent problem in spatial analyses of regional data. Chow and Lin (1971) (CL) were the first to develop a unified framework for the three problems (interpolation, extrapolation and distribution) of predicting disaggregated times series by so-called indicator series. This paper develops a spatial CL procedure for disaggregating cross-sectional spatial data and compares the Maximum Likelihood and Bayesian spatial CL forecasts with the naive *pro rata* error distribution. We outline the error covariance structure in a spatial context, derive the BLUE for the ML estimator and the Bayesian estimation procedure by MCMC. Finally we apply the procedure to European regional GDP data and discuss the disaggregation assumptions. For the evaluation of the spatial Chow-Lin procedure we assume that only NUTS 1 GDP is known and predict it at NUTS 2 by using employment and spatial information available at NUTS 2. The spatial neighborhood is defined by the inverse travel time by car in minutes. Finally, we present the forecast accuracy criteria comparing the predicted values with the actual observations.

1 INTRODUCTION

So far, the Chow and Lin (1971) method has been successfully applied to predict unobserved high-frequency time series data by related series. This paper will extend their paper in two directions: First, we will apply the procedure to regional cross-section data using a spatial econometric model (see Anselin (1988)) and second we will embed the model into a Bayesian framework. We address the problem of a regional data set that is completely observed at an aggregate level (NUTS 1) and has to be broken down into smaller regional units (NUTS 2) conditional on observables. We propose a spatial econometrics model in a Bayesian framework that has to be estimated by MCMC.

The paper is organized as follows. In section 2 we review the Chow-Lin method and discuss some regional disaggregation methods used in the literature recently. Section 3 presents the Maximum Likelihood model of the spatial CL method for cross-sectional data. The BLUE estimator for the spatial autoregressive model (SAR) is derived, along with the error covariance matrix needed for the prediction with the spatial gain term. We also outline the MCMC Bayesian extension to this approach. In section 4, we apply the spatial cross-sectional CL method to European NUTS 1 and NUTS 2 data. As we observe all data on the disaggregated level, we can compare the predicted values for GDP for NUTS 2 to their observed values and calculate forecast accuracy criteria. A final section concludes and gives an outlook for future extensions and research work.

2 THE CHOW-LIN METHOD

High frequency time series data of the economy is a valuable information for policy makers. However, such data are rarely available on a monthly or quarterly basis. In the past a lot of attempts have been made to interpolate missing high frequency data by using related series that are known. Friedman (1962) suggested relating the series in a linear regression framework. The three problems that are faced by the statisticians are known as interpolation, extrapolation and distribution problem of time series by related series. Interpolation is used to generate higher frequency stock data, while extrapolation extends given series outside the sample and distribution is used to add higher frequency observations to flow data such as GDP. The path-breaking paper by Chow and Lin (1971) extended the existing analysis to a unified framework, and led to a boost in research on this topic.

Assuming a linear relationship of the high frequency data $\mathbf{y} = \mathbf{X}\beta + \epsilon$, where $\mathbf{y}$ is a $(T \times 1)$ vector of unobserved high frequency data, $\mathbf{X}$ is a $(T \times k)$ matrix of observed regressors, β is a $(k \times 1)$ vector of coefficients and ϵ is a vector of random disturbances, with $E(\epsilon) = \mathbf{0}$ and $E(\epsilon\epsilon') = \sigma^2\mathbf{\Omega}$, Chow and Lin (1971) showed that the BLUE for the parameter vector $\hat{\beta}$ and the high frequency data $\hat{\mathbf{y}}$ is given by:

$$\hat{\beta} = (\mathbf{X}'\mathbf{C}'(\mathbf{C}\mathbf{\Omega}\mathbf{C}')^{-1}\mathbf{C}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}'(\mathbf{C}\mathbf{\Omega}\mathbf{C}')^{-1}\mathbf{C}\mathbf{y} \tag{1}$$

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} + \mathbf{\Omega}\mathbf{C}'(\mathbf{C}\mathbf{\Omega}\mathbf{C}')^{-1}(\mathbf{C}\mathbf{y} - \mathbf{C}\mathbf{X}\hat{\beta}) \tag{2}$$

Note that the matrix $\mathbf{C}$ aggregates the high frequency data to low frequency data. The essential part in the equations (1) and (2) is the residual covariance matrix $\mathbf{\Omega}$, which has to be estimated. So constructing the BLUE requires

knowledge or assumptions about the error covariance matrix. In the literature assumptions like random walk, white noise, random walk Markov or autoregressive process of order one have been suggested and tested (see Fernandez (1981), Di Fonzo (1990), Litterman (1983) or Pavia-Miralles et al. (2003) respectively). Some authors extended the framework for the multivariate case (see Rossi (1982) or Di Fonzo (1990)) covering time and space for example (see Pavia-Miralles and Cabrer-Borras (2007)). Usually constraints are imposed to restrict the predicted unobserved series to add up to the observed lower frequency series, e.g. by specifying penalty functions (see Denton (1971)). The discrepancy between the sum of the predicted high frequency observations and the corresponding low frequency observation is in this case divided upon the high frequency data through some assumptions (for example *pro rata* [1]).

There are various problems that may arise when applying the Chow-Lin (CL) procedure empirically. First, one has to find a suitable set of observable covariates. Predicted outcomes may heavily rely on the indicators chosen and their statistical properties. Seasonally adjusting the data and aggregating multicollinear variables improves the quality the results (see Pavia-Miralles and Cabrer-Borras (2007) for Monte Carlo evidence). Another crucial facet is of course the design of the residual covariance matrix and the restrictions imposed.

In recent years the CL procedure has been commonly used to construct unobserved monthly or quarterly data from observed annuals. However, no attempts have been made so far to apply the procedure to the geographical dimension. Regional econometric work in Europe has become increasingly important, especially since the integration process, the European Union puts a lot of weight on closing regional inequalities. For such analyses NUTS data are the main source of information. NUTS data are collected by the individual member states using common rules and methods. But not all member states have developed the same level and speed of skills, especially after 1995 when the harmonized European System of National and Regional Accounts started. This leads to inhomogeneous data quality and sometimes to gaps in the data base, especially of it comes to smaller regional units. In order to apply many modern panel methods, one has to complete such data sets. While the simplest method is trend interpolation, this gives not always satisfactory results and mostly the data isn't available at the required regional scale at all. Thus, there is a need for more elaborate methods, in particular since spatial econometric techniques have become popular in regional sciences.

LeSage and Pace (2004) use spatial econometric techniques to estimate missing dependent data. They predict unobserved house prices by using the informa-

[1] Pro rata means the error term will be distributed evenly among the high-frequency observations.

tion of sold and unsold houses to increase estimation efficiency. In contrast to what is implemented in this paper, LeSage and Pace (2004) predict unobserved spatially dependent data with observables at the same regional level. Our approach is more related to the CL procedure, where we observe the covariates at the disaggregated regional level and need to predict unobserved dependent variables at the same regional level conditional on the aggregate observed dependent variables. The spatial CL approach is therefor capable of predicting completely unobserved data.

3 DERIVATION OF THE SPATIAL CHOW-LIN FORECASTS

The spatial Chow-Lin procedure relies on 3 assumptions:

Assumption 1 *Structural similarity: The aggregated model for* $\mathbf{y}_c$ *and the disaggregated model for* $\mathbf{y}$ *are structurally similar. This implies that variable relationships that are observed on an aggregated level are following the same empirical law as on a disaggregated level: the regression parameters in both models are the same.*

Assumption 2 *Error similarity: The spatially correlated errors have a similar error structure on an aggregated level and on a disaggregated level: The spatial correlations are not significantly different.*

Assumption 3 *Reliable indicators: The indicators to make the formats on a disaggregated level have sufficiently large predictive power: The* R^2 *(or the F test) is significantly different from zero.*

Note that violation of assumption 1 leads to systematically biased forecasts. Violation of assumption 2 means that the spillover effects (the gain) in the CL prediction formula isn't substantially contributing to the forecasts. Violation of assumption 3 implies that we observe for the forecasts of the disaggregates only the simple proportions of the aggregates.

Consider a cross-sectional data set $\{\mathbf{y}, \mathbf{X}\}$ containing variables for N regions. Now assume that the unknown cross-sectional $N \times 1$ vector ($\mathbf{y}$), which is unobserved, satisfies a spatial autoregressive regression relationship (see e.g. Anselin 1988) of the following form:

$$\mathbf{y}_d = \rho_d \mathbf{W}_N \mathbf{y}_d + \mathbf{X}_d \beta_d + \epsilon_d, \qquad \epsilon_d \sim \mathcal{N}[\mathbf{0}, \sigma_d^2 \mathbf{I}_N]. \tag{3}$$

Here we adopt the notation that the disaggregated [2] vectors have a d subscript and matrices at the disaggregated regional scale have dimension N: $\mathbf{y}_d$ is the unknown $N \times 1$ dependent variable, and $\mathbf{X}_d$ is the known $N \times K$ disaggregated regressor matrix, which contains the so-called indicators that allow a good prediction of $\mathbf{y}$. ϵ_d is the error term in the disaggregated model and $\theta_d = (\beta_d, \rho_d, \sigma_d^2)$ are the disaggregated regression parameter.

The reduced form of the model (3) is obtained by the spread matrix $\mathbf{R}_N = \mathbf{I}_N - \rho_d \mathbf{W}_N$ for an appropriately chosen spatial weight matrix (see Anselin (1988)) at the disaggregated regional level $\mathbf{W}_N$:

$$\mathbf{y}_d = \mathbf{R}_N^{-1}\mathbf{X}_d\beta_d + \mathbf{R}_N^{-1}\epsilon_d, \quad \mathbf{R}_N^{-1}\epsilon_d \sim \mathcal{N}[\mathbf{0}, \mathbf{\Sigma}_d]. \tag{4}$$

with the disaggregate $N \times N$ covariance matrix

$$\mathbf{\Sigma}_d = \sigma_d^2(\mathbf{R}_N'\mathbf{R}_N)^{-1}, \tag{5}$$

that contains σ_d^2 and ρ_d. Now define a $n \times N$ matrix C consisting of 0's and 1's that aggregates the N subregions to n regions (with $N > n$) by pre-multiplication, like

$$\mathbf{y}_a = \mathbf{C}\mathbf{y}_d, \tag{6}$$

The C-aggregation of the reduced form model can thus be obtained by pre-multiplying model 4 by matrix $\mathbf{C}$, which leads to

$$\mathbf{C}\mathbf{y}_d = \mathbf{C}\mathbf{R}_N^{-1}\mathbf{X_d}\beta + \mathbf{C}\mathbf{R}_\mathbf{N}^{-\mathbf{1}}\epsilon, \quad \mathbf{C}\mathbf{R}_\mathbf{N}^{-\mathbf{1}}\epsilon \sim \mathcal{N}[\mathbf{0}, \mathbf{\Sigma_a}]. \tag{7}$$

with the aggregate covariance matrix $\mathbf{\Sigma}_a = \sigma_d^2\mathbf{C}(\mathbf{R}_N'\mathbf{R}_N)^{-1}\mathbf{C}'$. Note that the data for model (7) is observed, except for the disaggregate covariance matrix, which is needed to calculate the aggregated covariance matrix. As we cannot estimate the spatial lag parameter ρ_d for the unknown disaggregated vector $\mathbf{y}_d$ directly, we need another approach to get information about ρ_d. A simple estimate can be obtained by running the aggregated regression model:

$$\mathbf{y}_a = \rho_a\mathbf{W}_n\mathbf{y}_a + \mathbf{C}\mathbf{X}_d\beta + \nu_a, \quad \nu_a \sim \mathcal{N}[\mathbf{0}, \sigma_a^2\mathbf{I}_N], \tag{8}$$

[2] The term 'disaggregated' is used in analogy to the high-frequency data of the traditional time series CL procedure. The sum of a certain number of 'disaggregated' regions corresponds to what is throughout this paper denoted as 'aggregated' region.

using the steps described in LeSage (1998) for the 'mixed autoregressive' model. The aggregate estimate $\hat{\rho}_a$ can then be used to construct an estimate of the disaggregate spread matrix $\hat{\mathbf{R}}_N = \mathbf{I}_N - \hat{\rho}_a \mathbf{W}_N$. An estimate of the covariance matrix $\hat{\mathbf{\Sigma}}_d$ can then be derived from (5). Note that, by doing so, we assume that the spatial lag structure is the same at the disaggregate and aggregate regional level, which in fact is assumptions 1. However, to introduce the procedure of the spatial CL method, it is required that the assumptions 2 and 3 also hold.

The resulting GLS estimator for the spatial CL method can then be obtained by formulas similar to Chow and Lin (1971)

$$\hat{\beta}_{GLS} = (\mathbf{X}'\mathbf{C}'(\mathbf{C}\hat{\mathbf{\Sigma}}_d\mathbf{C}')^{-1}\mathbf{C}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}'(\mathbf{C}\hat{\mathbf{\Sigma}}_d\mathbf{C}')^{-1}\mathbf{y_a} \tag{9}$$

$$= (\mathbf{X}_a'(\mathbf{C}\hat{\mathbf{\Sigma}}_d\mathbf{C}')^{-1}\mathbf{X}_a)^{-1}\mathbf{X}_a'(\mathbf{C}\hat{\mathbf{\Sigma}}_d\mathbf{C}')^{-1}\mathbf{y_a}. \tag{10}$$

Note that $\hat{\beta}_{GLS}$ does not depend on σ_a^2 but on $\hat{\rho}_a$. Now a point forecasts for the disaggregated $\mathbf{y}_d$ can derived from

$$\hat{\mathbf{y}}_d = \hat{\mathbf{R}}_N^{-1}\mathbf{X_d}\hat{\beta}_{\mathbf{GLS}} + \hat{\mathbf{\Sigma}}_\mathbf{d}\mathbf{C}'(\mathbf{C}\hat{\mathbf{\Sigma}}_\mathbf{d}\mathbf{C}')^{-\mathbf{1}}(\mathbf{y_a} - \mathbf{C}\hat{\mathbf{R}}_\mathbf{N}^{-\mathbf{1}}\mathbf{X_a}\hat{\beta}_{\mathbf{GLS}}). \tag{11}$$

The first term on the left-hand side in (11), $\hat{\mathbf{y}}_0 = \hat{\mathbf{R}}_N^{-1}\mathbf{X_d}\hat{\beta}_{\mathbf{GLS}}$, is the naive prediction of the unknown vector $\mathbf{y}_d$. The second term consists of a spatial improvement of the Goldberger (1962) 'gain projection matrix'

$$\mathbf{G} = \hat{\mathbf{\Sigma}}_\mathbf{d}\mathbf{C}'(\mathbf{C}\hat{\mathbf{\Sigma}}_\mathbf{d}\mathbf{C}')^{-\mathbf{1}} \tag{12}$$

of dimension $N \times n$, which distributes the aggregated residual vector ν, $\hat{\nu} = \mathbf{y_a} - \mathbf{C}\hat{\mathbf{R}}_\mathbf{N}^{-\mathbf{1}}\mathbf{X_a}\hat{\beta_{\mathbf{GLS}}}$ to the disaggregated cross-sectional predictions of $\hat{\mathbf{y}}_d$. Thus, the Chow-Lin prediction vector can be written briefly as $\hat{\mathbf{y}}_d = \hat{\mathbf{y}}_0 + Gain * Residual$ or $\hat{\mathbf{y}}_d = \hat{\mathbf{y}}_0 + \mathbf{G}\hat{\nu}$.

The gain crucially depends on the estimated spatial lag parameter ρ_a of the aggregate model. Note that if $\rho = 0$ then $\mathbf{\Sigma}_d = \mathbf{I}_N$ and the gain matrix reduces to a transposed projection matrix: $\mathbf{G} = \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}$ which amounts to an 'inverse averaging' or 'distributing matrix' since $\mathbf{C}\mathbf{C}'$ gives a diagonal matrix with the number of subregions for each aggregated region. Thus, the gain component $\mathbf{G}\hat{\nu}$ is a vector which in the $\rho = 0$ case is a down-weighted aggregated residual of the aggregated fit: A large residual will be smoothed out over N subregions and $1/N$-th is added to the simple forecasts. For example, if a certain region has a residual that lies below average, then all the disaggregated forecasts with gains will have their simple forecasts corrected downward. The same will

happen in the other direction, when the aggregated residual is positive. The effect of the spatial ρ is a 'spatial smearing out' of these $1/N$ discounted aggregated residual to the spatial neighbors. Thus, these point forecasts for the disaggregated model are called 'with spatial gain' in this paper, opposed to the models without them ($\rho = 0$) which are called naive predictions. One of the main interests for applied research is, whether the spatial gain prediction of $\mathbf{y}_d$ outperforms the naive prediction by means of forecasting criteria.

Proof 1 *The proof of the (reduced form) CL forecasting formula relies on the fact that in the joint distribution, the aggregated and the disaggregated model are correlated:*

$$p\begin{pmatrix}\mathbf{y}_d\\ \mathbf{y}_a\end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix}\mathbf{X}_a\beta\\ \mathbf{X}_c\beta\end{pmatrix}, \begin{pmatrix}\mathbf{\Sigma}_d & \mathbf{\Sigma}_d\mathbf{C}'\\ \mathbf{C}\mathbf{\Sigma}_d & \mathbf{C}\mathbf{\Sigma}_d C'\end{pmatrix}\right]. \tag{13}$$

Then, using the formula for the conditional normal distribution, we derive the conditional predictive distribution for the disaggregates

$$p(\mathbf{y}_d \mid \mathbf{y}_a, ...) = \mathcal{N}\left[\mu_d, \mathbf{\Sigma}_d \mid \mathbf{y}_a, \right]$$

which will further be used in Theorem (2).

3.1 BAYESIAN ESTIMATION

This section outlines the Bayesian implementation of the spatial CL procedure. The prior distribution for the parameters $\theta_a = (\beta_{\mathbf{a}}, \sigma_{\mathbf{a}}^{-\mathbf{2}}, \rho_{\mathbf{a}})$ is proportional to

$$\begin{aligned} p(\theta_a) &\propto p(\beta_a) \cdot p(\sigma_a^{-2}) && (14)\\ &= \mathcal{N}[\beta_a \mid \mathbf{b}_*, \mathbf{H}_*] \cdot \mathbf{\Gamma}(\sigma_{\mathbf{a}}^{-\mathbf{2}} \mid \mathbf{s}_*^{\mathbf{2}}, \mathbf{n}_*), && (15)\end{aligned}$$

where all parameters with subscript $*$ denote known hyper-parameters, since we assume a uniform prior for $\rho_a \sim U[-1, 1]$. As before, the C-aggregation of the reduced form model is obtained by multiplying with the $N \times n$ matrix $\mathbf{C}$

$$\mathbf{y}_a = \mathbf{C}\mathbf{R}_N^{-1}\mathbf{X}_{\mathbf{d}}\beta_a + \mathbf{C}\mathbf{R}_N^{-1}\epsilon, \quad \mathbf{C}\mathbf{R}_N^{-1}\epsilon \sim \mathcal{N}[\mathbf{0}, \mathbf{C}\mathbf{\Sigma}_d\mathbf{C}']. \tag{16}$$

In the following we make the convention that if we omit the index for a parameters we deal with aggregates, since estimation can only be done for aggregates. The joint distribution of $\theta_a = (\beta, \rho, \sigma^2)$ of this model is given by

$$p(\theta \mid \mathbf{y}_a) = \mathcal{N}[\mathbf{C}\mathbf{R}_N^{-1}\mathbf{X}\beta, \mathbf{C}\boldsymbol{\Sigma}_d\mathbf{C}']\mathcal{N}[\beta \mid \mathbf{b}_*, \mathbf{H}_*] \cdot \boldsymbol{\Gamma}(\sigma^{-\mathbf{2}} \mid \mathbf{s}_*^{\mathbf{2}}, \mathbf{n}_*) \tag{17}$$

Consider the SAR cross-sectional CL model and let us denote the 3 conditional distributions by $p(\rho \mid \theta^c), p(\beta \mid \theta^c)$, and $p(\sigma^2 \mid \theta^c)$ where $\theta = (\rho, \beta, \sigma^2)$ denotes all the parameters of the model and θ^c the complementary parameters in the f.c.d.'s, respectively. The MCMC procedure consists of 3 blocks of sampling, as is shown in the following theorem:

Theorem 1 (MCMC in the SAR-CL model) *Repeat the following steps until convergence:*

(1) Draw β from $\mathcal{N}[\beta \mid \mathbf{b}_{**}, \mathbf{H}_{**}]$
(2) Draw ρ_i by a Metropolis step: $\rho_{new} = \rho_{old} + \mathcal{N}(0, \tau^2)$
(3) Draw σ^{-2} from $\Gamma[\sigma^{-2} \mid s_{**}^2, n_{**}]$

Proof 2 *The fcd for the aggregated β regression coefficients is*

$$\begin{aligned} p(\beta \mid \mathbf{y_a}, \theta^{\mathbf{c}}) &= \mathcal{N}[\beta \mid \mathbf{b}_*, \mathbf{H}_*] \cdot \mathcal{N}[\mathbf{y_a} \mid \mathbf{C}\mathbf{R}_{\mathbf{N}}^{-\mathbf{1}}\mathbf{X}\beta, \mathbf{C}\boldsymbol{\Sigma}_{\mathbf{d}}\mathbf{C}'] && (18) \\ &= \mathcal{N}[\beta \mid \mathbf{b}_{**}, \mathbf{H}_{**}] && (19) \end{aligned}$$

with the parameters

$$\begin{aligned} \mathbf{H}_{**}^{-\mathbf{1}} &= \mathbf{H}_*^{-\mathbf{1}}\mathbf{b}_* + \sigma^{-\mathbf{2}}\mathbf{X}_{\mathbf{a}}'\mathbf{R}_{\mathbf{N}}'^{-\mathbf{1}}\mathbf{C}'\boldsymbol{\Sigma}_{\mathbf{d}}^{-\mathbf{1}}\mathbf{C}\mathbf{R}_{\mathbf{N}}^{-\mathbf{1}}\mathbf{X}_{\mathbf{a}}, && (20) \\ \mathbf{b}_{**} &= \mathbf{H}_{**}[\mathbf{H}_*^{-\mathbf{1}}\mathbf{b}_* + \sigma^{-\mathbf{2}}\mathbf{X}_{\mathbf{a}}'\mathbf{R}_{\mathbf{N}}'^{-\mathbf{1}}\mathbf{C}'\boldsymbol{\Sigma}_{\mathbf{d}}^{-\mathbf{1}}\mathbf{y}]. && (21) \end{aligned}$$

Proof 3 *The fcd for the residual variance we find*

$$p(\sigma^{-2} \mid \mathbf{y}_a, \theta^c) = \Gamma[\sigma^{-2} \mid s_{**}^2, n_{**}] \tag{22}$$

with $n_{**} = n_* + n$ and $s_{**}^2 n_{**} = s_*^2 n_* + ESS$ where the error sum of squares ESS is given by

$$ESS = (\mathbf{y}_a - \mathbf{C}\mathbf{R}^{-1}\mathbf{X}\beta)'\mathbf{C}\boldsymbol{\Sigma}_d^{-1}\mathbf{C}'(\mathbf{y}_a - \mathbf{C}\mathbf{R}^{-1}\mathbf{X}\beta). \tag{23}$$

Proof 4 *For drawing the ρ's we use a Metropolis step:*

$$\rho_{new} = \rho_{old} + N(0, \tau^2)$$

with the acceptance ratio

$$\alpha = min\left[1, \frac{p(\rho_{new})}{p(\rho_{old})}\right]$$

where $p(\rho)$ is the (kernel of) the full conditional for ρ, in our case the kernel is just stemming from the likelihood function:

$$p(\rho) = |\mathbf{\Omega}(\rho)|^{-\frac{1}{2}} exp(-\frac{1}{\sigma^2} ESS_\rho) \tag{24}$$

with ESS_ρ given in (23).

From the MCMC simulation we obtain a numerical sample of the posterior distribution $p(\beta, \rho, \sigma^{-2} \mid \mathbf{y})$. Based on this sample we obtain the posterior predictive distribution by integrating over the conditional predictive distribution with the posterior distribution

$$p(\mathbf{y}_p \mid \mathbf{y}) = \int \mathbf{p}(\mathbf{y_p} \mid \theta)\mathbf{p}(\theta, \mid \mathbf{y_a})\mathbf{d}\theta \tag{25}$$

where the posterior density for θ, $p(\beta, \rho, \sigma^{-2} \mid \mathbf{y})$ is given numerically by the MCMC sample. In case the posterior distribution $p(\theta \mid \mathbf{y})$ is not given in closed form but is available as a MCMC sample, the integration over θ in (25) is done by direct simulation to obtain a predictive density in form of a sample $\mathbf{Y}_p = \{\mathbf{y}_p^{(1)}, ..., \mathbf{y}_p^{(J)}\}$.

Let us denote the posterior sample of the θ_a parameters from the MCMC procedure by:

$$\mathbf{\Theta}_{MCMC} = \{(\beta_j, \rho_j, \sigma_j^2), \quad j = 1, ..., J\}. \tag{26}$$

From this output we find a predictive sample of the unknown disaggregated $N \times 1$ vector $\mathbf{y}$ by drawing from the reduced form (which depends on the disaggregated matrix $\mathbf{W}_N$ and on the known disaggregated regressors $\mathbf{X}_d$). This amounts to the task of determining the predictive distribution of a regression model, where the conditional predictive distribution is now taken from the disaggregated model.

We summarize this procedure in the next theorem:

Theorem 2 (Bayesian prediction in the SAR-CL model) *From the joint*

distribution of the aggregated and the disaggregated model we find for the conditional distribution of the disaggregated dependent variable $\mathbf{y}_d$*:*

$$\mathbf{y}_d^{(j)} \sim \mathcal{N}[\mathbf{R}_{Nj}^{-1}\mathbf{X}\beta_j + \mathbf{g}_j, \sigma_j^2[(\mathbf{R}'_{Nj}\mathbf{R}_{Nj})^{-1} - \mathbf{G}_j]], \tag{27}$$

with $\mathbf{R}_{Nj} = (\mathbf{I}_N - \rho_{nj}\mathbf{W}_N), j = 1, ..., J$ *and* $\mathbf{g}_j$ *is the gain vector that shifts the conditional mean*

$$\mathbf{g_j} = (\mathbf{R}'_{\mathbf{Nj}}\mathbf{R}_{\mathbf{Nj}})^{-\mathbf{1}}\mathbf{C}'[\mathbf{C}(\mathbf{R}'_{\mathbf{Nj}}\mathbf{R}_{\mathbf{Nj}})^{-\mathbf{1}}\mathbf{C}']^{-\mathbf{1}}(\mathbf{y_a} - \hat{\mathbf{y}}_{\mathbf{aj}})], \tag{28}$$

and $\mathbf{G}_j$ *is the (positive definite) gain matrix for the covariance matrix*

$$\mathbf{G}_j = (\mathbf{R}'_{Nj}\mathbf{R}_{Nj})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{R}'_{Nj}\mathbf{R}_{Nj})^{-1}\mathbf{C}']^{-1}\mathbf{C}(\mathbf{R}'_{Nj}\mathbf{R}_{Nj})^{-1}], \tag{29}$$

which reduces the unconditional covariance matrix $(\mathbf{R}'_{Nj}\mathbf{R}_{Nj})^{-1}$ *in (27).*

Proof 5 *Using the formula for conditional normal distributions for partitioned multivariate normal distributions as in (13). The drawing procedure is called direct sampling in Geweke (2005).*

We can summarize the procedure in the following 'prediction box':

SAR-CL prediction:

(1) For each $\theta_j \in \mathbf{\Theta}_{MCMC}, \quad j = 1, ..., J$
(2) Draw $\mathbf{y}^{(j)}$ from the CL conditional density $\mathcal{N}\left[\ \mu_d^{(j)}, \Sigma_d^{(j)}\right]$

with

$$\mu_d^{(j)} = \mathbf{R}_{\mathbf{Nj}}^{-\mathbf{1}}\mathbf{X_d}\beta_\mathbf{j} + \mathbf{g_j}$$

and

$$\mathbf{\Sigma}_d^{(j)} = \sigma_j^2[(\mathbf{R}'_{Nj}\mathbf{R}_{Nj})^{-1} - \mathbf{G}_j]$$

Note that for the Bayesian CL prediction of the disaggregated values $\mathbf{y}_d$ the 'CL conditional density' is a double conditional distribution, since we do not only condition on the aggregated parameters θ_a but also on the aggregated observations, the dependent variable $\mathbf{y}_a$. In the numerical simulation, we just integrate over the aggregated θ_a, but keep the dependency on the aggregated observations $\mathbf{y}_a$, since this should be one source of improved forecasts.

4 APPLICATION

This section presents an application of the developed spatial Chow-Lin method. We apply the Maximum Likelihood and Bayesian CL approach to European regional data and evaluate the resulting series against a naive (non-spatial) error distribution approach. In the naive prediction the disaggregated covariance matrix is the identity matrix ($\Sigma_N = I_N$).

In doing so, the resulting gain matrix just divides the estimated aggregated residuals evenly among the subregions. We then compare the predictions of the spatial CL and the naive spatial prediction to the observed series and evaluate their prediction accuracy by conventional forecasting criteria.

4.1 Data

We use data on regional GDP and employment of 2005 for 239 NUTS 2 regions of the EU27 from the Eurostat Regional Database. [3] By pre-multiplication of a constructed C-matrix, the 239 NUTS 2 regions aggregate up to 84 NUTS 1 regions.

The total number of NUTS 2/NUTS 1 regions corresponds to the following distribution among countries: Austria (9/3), Belgium (11/3), Bulgaria (6/2), Cyprus (1/1), Czech Republic (8/1), Germany (36/13), Estonia (1), Spain (19/7), Finland (5/2), France (22/8), Greece (9/3), Hungary (7/3), Ireland (2/1), Italy (22/5), Lithuania (1/1), Luxembourg (1/1), Latvia (1/1), Malta (1/1), Netherlands (12/4), Poland (16/6), Portugal (5/1), Romania (8/4), Slovakia (2/1) and the UK (33/11).

As a spatial weight matrix we use inverse car travel times between the capitol cities of the NUTS 2 regions, obtained from map24. [4] Due to this derivation of the travel times all islands, autonomous and extraterritorial regions are excluded from the sample. Additionally, we excluded all regions that are not conform with the old classification system (2004) of the NUTS codes (Sweden, Slovenia and Denmark completely and some regions of Germany). For the construction of the spatial weight matrix W we chose a neighboring ratio cut-off point of 180 minutes. This means that we only regard the neighborhood of regions within a radius of 180 minutes, the rest is set to zero. For the sensitivity analysis we use thresholds of 90 and 360 minutes. We then take the inverse of the resulting matrix and normalize the row sums to one. The

[3] http://epp.eurostat.ec.europa.eu/
[4] http://www.at.map24.com/

aggregate travel time matrix is calculated using the arithmetic means of the disaggregated regions to each other.

Table 1 displays the results for the two regional levels NUTS 1 and NUTS 2 for Moran's I test using the inverse travel time by car as a spatial weight matrix and employment and an intercept as explanatory variables. Note that in this experimental situation we observe the disaggregated data and can check the spatial properties of the series.

Table 1
Moran's I of NUTS 1 and NUTS 2 data

	NUTS 1	NUTS 2
Moran's I	0.6654	0.3312
Moran's I statistic	3.1078	7.6106
Marginal probability	0.0019	0
Number of obs.	84	239
Number of variables	2	2

The residuals exhibit significant spatial correlation. With a Moran's I of 3.1 and 7.6 we can reject the null hypothesis of no spatial correlation in the residuals. A spatial regression model should therefore correct part of the bias of the OLS estimation. Table 1 also provides first evidence in favor of assumption 2, as we observe a spatial correlation on both regional scales.

4.2 Estimation of the aggregate models

In order to derive the GLS ML estimate of β from equation 9 and the spatial CL prediction from equation 11, an estimate of ρ is needed, to construct the disaggregate covariance matrix from equation 5. For the Bayesian method, one requires a MCMC sample of the parameters (see theorem 1). So the first step is the Maximum Likelihood and Bayesian estimation of the SAR models. The dependent variable of our models is NUTS 1 GDP in nominal terms of 2005, the independent variables are an intercept (c), total employment (emp) in 2005 and the spatial lag of GDP ($GDP * W^{90}$) using the inverse car travel times up to a cut-off point of 90 minutes. [5] The estimation results [6] are presented in the following table.

The second and third column in table 2 display the results for the aggregate regression. For both estimation methods, the Maximum Likelihood and

[5] The superscript of the spatial weight matrix indicates the threshold in minutes
[6] The SAR and SAR_g program of LeSage (1997) spatial econometrics toolbox have been used to obtain the estimates.

Table 2
Model 1 - basic model regression results, Dependent Variable GDP, 2005

	NUTS 1			NUTS 2			
Variable	ML		Bayesian	ML		Bayesian	
c	-15131.14	***	-11602.45	-22680.26	***	-13442.97	***
	(3585.58)		(3674.09)	2969,40		(2172.2)	
emp	50.82	***	49.71	60.04	***	52.9	***
	(4.24)		(4.26)	(2.2)		(1.98)	
$GDP * W^{180}$	0.36	***	0.32	0.25	***	0.19	***
	(0.05)		(0.05)	(0.04)		(0.03)	
R^2	0.7223		0.7288	0.7262		0.7271	
Observations	84		84	239		239	

***(**)[*] indicates that the coefficient is significant at a 1% (5%) [10%] level.

Bayesian, all estimated coefficients are highly significant and show the expected sign. Higher employment stimulates GDP and the spatial lag parameter, describing the neighborhood relationship within a 180 minutes car travel radius, is positive and statistically significant with a value of 0.36. Given the assumptions in section 3 hold, we should receive an adequate forecasts as also a high R^2 of around 0.7 is observed.

We can further check the assumptions of the CL procedure by estimating the disaggregated regression model [7]. A summary of the results is presented in the last two columns of table 2. The signs, significance levels and the R^2 correspond to the aggregate regression. The coefficients are somehow different in size, as the spatial lag parameter and the intercept are smaller, whereas the the coefficient of employment is larger.

4.3 Prediction accuracy

To assess the prediction accuracy of our two approaches, we first check the correlation of the NUTS 2 CL predicted data and the observed data series. This assessment is done for the ML and Bayesian prediction as well as for the CL procedure with and without the spatial gain term. From table 3 we see that in both cases, for the Bayesian and the ML prediction, the procedures with

[7] Recall that this is usually not possible as the disaggregated dependent variable is not available in practice.

the spatial gain term outperform the naive CL *pro rata* error distributions, in which the error term is distributed evenly among the subregions. This is an indication that the spatial gain term, in general, does improve the forecasts on a disaggregated level. It can also be observed that the Bayesian prediction yields better fits of the observed data, whereas the difference between the naive and the spatial gain forecast nearly vanishes. This may be due to the fact that there is no restriction included, guaranteeing the equality between the sum of the predicted NUTS 2 values and the observed NUTS 1 values.

Table 3
Correlation between naive or spatial CL prediction with observed data (NUTS 2)

	naive CL	CL with spatial gain
ML	0.953	0.963
Bayesian	0.962	0.964

To numerically evaluate the accuracy of the ML and Bayesian prediction we chose three measures from the forecasting literature (see Chatfield (2001)), the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE) [8] .

The CL prediction accuracy measures are shown in table 7 of the annex. Within the Maximum Likelihood framework, it can be seen that there are clear improvements according to all prediction accuracy criteria when switching from the naive to the spatial error distribution. Turning to the comparison within the Bayesian method, the results stay the same. However, when comparing the ML to the Bayesian approach, it seems that the Bayesian method generates more precise forecasts.

4.4 Sensitivity analysis

As a sensitivity check, we alter the car travel time threshold to 90 and 360 minutes, reestimate the equations and calculate the new prediction criteria. Table 4 of the annex shows the estimation results for the aggregate and disaggregate model with a travel time threshold of 90 minutes (model 2). The estimate of the disaggregate ρ deviates more than twice its standard deviation from the point estimate of the aggregate regression. This violates assumptions 1 and 2, further leading to biased predictions of the disaggregate series. The impact on the CL predictions can be recoverd from the second panel of table 7 of the annex. We see that Bayesian and ML naive predictions account for smaller forecasting errors than their spatial CL counterparts. So the violation

[8] The formulas are $RMSE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(y-\widehat{y})^2}$, $MAE = \frac{1}{N}\sum_{i=1}^{N}|y-\widehat{y}|$ and $MAPE = \frac{1}{N}\sum_{i=1}^{N}|\frac{y-\widehat{y}}{y}|$ respectivley.

of assumption 2 - error similarity - of the spatial CL procedure, led to worse forecasts when compared to the naive *pro rata* error distribution.

By contrast, table 5 reveals rather similar spatial lag coefficients at the disaggregate and aggregate regional scale for the threshold of 360 minutes (model 3). In the third panel of table 7 the forecast criteria for the predictions of the model with a thresholds of 360 minutes is summarized. Because of the similar structural relationship and error structure, the CL predictions for both ML and Bayesian estimations generate more precise predictions.

Finally, we add additional regressors to the model, to check the properties of an improvement in R^2 on the prediction accuracy (see assumption 3). This model (model 4) particularly explores the possibility that there is more spatial variation in the regressors and tests whether there are two regimes - Eastern and Western Europe. In particular, we augment the model by an eastern European dummy (dum_{east}) that equals zero for regions of the old EU-15 member states and one for regions of the 12 new members after 2004, a spatial durbin lag of employment ($emp * W^{90}$) that can be interpreted as 'potential variable' and interaction terms of the eastern European dummy with employment ($dum_{east} * emp$) and the spatial durbin lag ($dum_{east} * emp * W^{90}$). A summary of the estimation results is given in table 6 of the annex.

The R^2 increases to 0.91 and 0.86 for the aggregate and disaggregate model respectively. For both aggregation levels and estimation methods we find significant impacts of employment (smaller impact for Eastern Eurpean regions) and a positive significant spatial lag. However, on the aggregate level (first two columns in the table) we notice large differences in the coefficients for the spatial durbin lag of employment and the spatial lag parameter between the ML and the Bayesian estimation method. At the disaggregate level, the spatial durbin lag for Eastern Europe only effects GDP in the Bayesian framework, whereas the coefficient of the eastern Europe dummy is three times as high in the ML estimation, though not significant. Even though the Eastern European dummy itself does not turn out significant, the results for the influence of employment point to regional heterogeneity between Eastern and Western European countries.

The corresponding prediction accuracy measures for model 4 are displayed in the fourth panel of table 7. Within the extended model, the spatial gain again improves the forecasts, most likely due to the similar spatial error structure identified on the aggregate and disaggregate regional scale. As before the predictions of the Bayesian estimation and prediction method account for lower forecasting errors. Based on the basic model (model 1), the forecasts of the extended model (model 4) exhibit lower forecasting errors. The improvement in R^2 led to better forecasts (see assumption 3). This example showed that it pays to explore spatial heterogeneity in European regions. Once a reliable

model can be found, the reward is a better CL forecasting ability for the disaggregated regions.

5 CONCLUSIONS AND OUTLOOK

This paper has shown that the Chow-Lin method can be applied to regional cross-section data, where the different NUTS classifications represent the high and low frequency data. Given that the assumption of structural similarity, error process similarity and reliability of the indicators hold, the procedure should improve the predictions compared to naive *pro rata* error distribution. Besides the Maximum Likelihood estimator, we present a Bayesian estimator, that has the advantage that the problem of missing data can be easily implemented within the widely available MCMC programs.

Application results for European GDP data on NUTS 2 and 1 level showed the advantages of using the spatial Chow-Lin approach vis-à-vis the naive non-spatial *pro rata* distribution of the error terms. This is illustrated using conventional forecasting criteria as well as the simple correlations between the predictions and the observed data. We find that if the three assumptions hold, the spatial Chow-Lin procedure clearly outperforms the naive Chow-Lin error distribution, whereas the Bayesian method further improves the prediction compared to the Maximum Likelihood estimation.

It is further planned to develop the method for panel data and panel flow models with spatial patterns, extending the approach of this paper. In further studies, we will explore the possibility to use spatial CL methods for even smaller units like NUTS 3 (Polasek et al. (2009)). Besides predicting level data, transformations like logarithms, per capita values and growth rates may be of interest to the researcher. Apart from that, the composition of the related variables is crucial for the prediction accuracy of the disaggregate series. In our example, we just used the simple relationship between employment and GDP, as rich data sets on regional level are hardly available. However, the spatial CL method seems to be a reasonable tool to deal with the problems of incomplete or unobserved regional data.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Chatfield, C. (2001). *Time-series Forecasting*. Chapman & Hall.

Chow, G. C. and Lin, A. (1971). Best linear unbiased interpolation, distri-

bution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53(4):372–375.

Denton, F. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of American Statistical Association*, 66:99–102.

Di Fonzo, T. (1990). The estimation of m disaggregate time series when contemporaneous and temporal aggregates are known. *The Review of Economics and Statistics*, 71:178–182.

Fernandez, R. B. (1981). A methodological note on the estimation of time series. *The Review of Economics and Statistics*, 53:471–478.

Friedman, M. (1962). The interpolation of time series by related series. *Journal of American Statistical Association*, 57:729–757.

Geweke, J. F. (2005). *Contemporary Bayesian econometrics and statistics*. Wiley series in probability and statistics.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *American Statistical Association Journal*, 57:369–375.

LeSage, J. P. (1997). Bayesian estimation of spatial autoregressive models. *International Regional Science Review*, 20:113–129.

LeSage, J. P. (1998). Spatial econometrics. Technical report, University of Toledo.

LeSage, J. P. and Pace, R. K. (2004). Models for spatially dependent missing data. *Journal of Real Estate Finance and Economics*, 29:233–254.

Litterman, R. B. (1983). A random walk, markov model for the distribution of time series. *Journal of Business and Economic Statistics*, 1:169–173.

Pavia-Miralles, J. M. and Cabrer-Borras, B. (2007). On estimating contemporaneous quarterly regional gdp. *Journal of Forecasting*, 26:155–170.

Pavia-Miralles, J. M., Vila-Lladosa, L.-E., and Valles, R. E. (2003). On the performance of the chow-lin procedure for quarterly interpolation of annual data: Some monte-carlo analysis. *Spanish Economic Review*, 5:291–305.

Polasek, W., Sellner, R., and Llano, C. (2009). Interpolating spanish nuts 3 data by spatial chow-lin. forthcoming.

Rossi, N. (1982). A note on the estimation of disaggregate time series when the aggregate is known. *Review of Economics and Statistics*, 64:695–696.

6 ANNEX

Table 4
Model 2 - spatial threshold = 90 minutes, dependent variable: GDP NUTS 1 and 2, 2005

	NUTS 1				NUTS 2			
Variable	ML		Bayesian		ML		Bayesian	
c	-10535.98	***	-5320.00	**	-14886.52	***	-6962.17	***
	(4141.5)		(3616.01)		(2736.49)		(1899.39)	
emp	54.42	***	48.11	***	59.57	***	52.81	***
	(4.74)		(4.46)		(2.32)		(2.21)	
$GDP * W^{90}$	0.23	***	0.26	***	0.12	***	0.09	***
	(0.05)		(0.06)		(0.04)		(0.03)	
R^2	0.7482		0.7408		0.7405		0.7312	
Observations	84		84		239		239	

***(**)[*] indicates that the coefficient is significant at a 1% (5%) [10%] level.

Table 5
Model 3 - spatial threshold = 360 minutes, dependent variable: GDP NUTS 1 and 2, 2005

	NUTS 1				NUTS 2			
Variable	ML		Bayesian		ML		Bayesian	
c	-19278.91	***	-15708.02	***	-31001.52	***	-21464.97	***
	(3825.18)		(3976.16)		(3166.65)		(2526.59)	
emp	54.62	***	53.32	***	59.66	***	51.82	***
	(4.25)		(4.37)		(2.12)		(2.01)	
$GDP * W^{360}$	0.37	***	0.32	***	0.43	***	0.37	***
	(0.06)		(0.06)		(0.05)		(0.04)	
R^2	0.6820		0.6993		0.6917		0.7033	
Observations	84		84		239		239	

***(**)[*] indicates that the coefficient is significant at a 1% (5%) [10%] level.

Table 6
Model 4 - Extended specification, dependent variable: GDP NUTS 1 and 2, 2005

	NUTS 1				NUTS 2			
Variable	ML		Bayesian		ML		Bayesian	
c	-4157.94		-4535.06	*	-7746.00	***	-2485.20	**
	(3207.84)		(2751.15)		(3166.74)		(1293.86)	
dum_{east}	5200.94		4900.15		5823.59		1612.85	
	(8042.1)		(5634.86)		(7190.49)		(2758.27)	
emp	59.383	***	55.59	***	64.21	***	56.50	***
	(2.77)		(2.96)		(1.72)		(1.42)	
$dum_{east} * emp$	-49.45	***	-45.01	***	-49.05	***	-43.60	***
	(6.98)		(5.4)		(5.74)		(2.87)	
$emp * W^{90}$	-11.89	***	-3.72		-10.69	*	-10.19	***
	(3.13)		(3.52)		(5.71)		(2.51)	
$dum_{east} * emp * W^{90}$	10.85		4.10		7.83		8.03	***
	(8.51)		(6.35)		(6.65)		(2.59)	
$GDP * W^{90}$	0.29	***	0.20	***	0.20	**	0.18	***
	(0.04)		(0.04)		(0.08)		(0.04)	
R^2	0.9184		0.9129		0.8638		0.8510	
Observations	84		84		239		239	

***(**)[*] indicates that the coefficient is significant at a 1% (5%) [10%] level.

Table 7
Prediction accuracy

Model 1 - Basic Model				
		RMSE	MAE	MAPE
ML	naive CL	16001	9794	0.867
	CL with spatial gain	14072	8720	0.678
Bayesian	naive CL	13761	8644	0.605
	CL with spatial gain	13405	8331	0.519
Model 2 - Spatial neighborhood threshold = 90 minutes				
		RMSE	MAE	MAPE
ML	naive CL	14997	8982	0.656
	CL with spatial gain	16159	9414	0.664
Bayesian	naive CL	13982	8398	0.502
	CL with spatial gain	16966	9437	0.526
Model 3 - Spatial neighborhood threshold = 360 minutes				
		RMSE	MAE	MAPE
ML	naive CL	14050	7798	0.698
	CL with spatial gain	12876	7553	0.625
Bayesian	naive CL	12323	7193	0.541
	CL with spatial gain	12081	7259	0.518
Model 4 - Extended Specification				
		RMSE	MAE	MAPE
ML	naive CL	15503	8654	0.737
	CL with spatial gain	13821	7835	0.577
Bayesian	naive CL	11501	6153	0.391
	CL with spatial gain	11241	6052	0.350